E7.5-10.2.9.5
CR-142839

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
P. O. BOX 618 • ANN ARBOR • MICHIGAN • 48107
PHONE (313) 483-0500

Information Systems and Analysis

101900-55-L
21 May 1975

Developing Processing Techniques for Skylab Data
Monthly Progress Report, April 1975

(E75-10295) DEVELOPING PROCESSING TECHNIQUES FOR SKYLAB DATA Monthly Progress Report, Apr. 1975 (Environmental Research Inst. of Michigan) 12 p HC $3.25 CSCL 09F

N75-25242
Unclas
G3/43 00295

EREP Investigation 456 M
NASA Contract NAS9-13280

Prepared by

Richard F. Nalepka - Principal Investigator
William A. Malila - Co-Principal Investigator
James P. Morgenstern

NASA Technical Monitor

Mr. Larry B. York/TF6
National Aeronautics and Space Administration
Johnson Space Center
Principal Investigator Management Office
Houston, Texas 77058

Developing Processing Techniques for Skylab Data
Monthly Progress Report, April 1975

The following report serves as the twenty-sixth monthly progress report for EREP Investigation 456 M which is entitled "Developing Processing Techniques for Skylab Data". The financial report for this contract (NAS9-13280) is being submitted under separate cover.

The purpose of this investigation is to test information extraction techniques for SKYLAB S-192 data and compare with results obtained in applying these techniques to LANDSAT and aircraft scanner data.

In previous reports we had considered the question of SDO-SDO spatial misregistration of the SKYLAB S-192 multispectral scanner. We had also reported on use of an automated technique for locating fields of interest.

During the reporting period we completed one phase, an analysis of the conic data, of the spatial misregistration study outlined in the previous report. We began a second investigation concerning misregistration, this into the effects of misregistration on classification and acreage estimation accuracy. We also extracted signatures for the primary ground covers in the test area. These will form the basis for a series of signature analyses, as outlined below.

DETERMINATION OF SPATIAL MISREGISTRATION

The previous monthly report described a method for determining the amount of misregistration between two correlated data channels. The algorithm described in that report has since been programmed, debugged and tested on conical Skylab data taken from the Michigan test site.

Initial tests indicated that the misregistration estimate was being biased by the DC (average) component of the signal in each channel. To remove this bias, the algorithm was modified to subtract out the mean value of each channel before computing the cross correlation. In essence, the cross correlation between the AC (varying) components of the signals are now being computed. This modification removed the bias that was noted.

To determine the misregistration between two channels, the cross correlation is determined over a range of fractional pixel shifts. The cross correlation peaks near the shift representing the actual misregistration and slopes down on both sides of this peak. Initial tests of the method indicated

that the values near the peak closely approximated a quadratic curve. To obtain a more accurate estimate of the shift at which the peak actually occurs, a quadratic curve was fitted to the three shift values nearest the peak. From the coefficients of this curve, the peak of the cross-correlation function can be easily estimated. In this manner, the peak of the cross-correlation can be estimated as a value within the fractional pixel shifts for which the function is actually determined.

Table I contains the estimated misregistration between 17 of the original 22 Skylab SDO's. Two of the SDO's (15,16) not appearing in the table, are not being used in the current Skylab investigation. The remaining three SDO's not in the table (18,21,22), were not sufficiently correlated with any other channels to obtain meaningful results.

The misregistration was not actually determined by direct measurement for all of the pairs of channels represented in the table. The misregistration was first measured between seven pairs of even and odd numbered high sample rate SDO's (1-2, 3-4, 5-6, 7-8, 9-10, 11-12, 13-14). In all cases, the average measurement taken over 5 lines of data was almost exactly 0.5. These measurements indicated that the misregistration between these pairs of channels could be safely assumed as being 1/2 pixel. Measurements were made using 10 lines of conical data on an additional seventeen pairs of correlated ($\rho \geq .5$ for a large sample of pixels) channels chosen from among the odd numbered high sample rate channels and the remaining low sample rate channels. A multiple linear regression was performed on these seventeen measurements to obtain estimates of the misregistration between nine pairs of channels from which estimates of all of the remaining pairs were derived. The sum of the squared deviations between the 17 actual measurements and their predicted values from the regression analysis was 0.0015. This low figure indicates the consistency of the results obtained from the different pairs of channels. As a further test, measurements of the misregistration between nine pairs of channels taken from a different set of 10 lines, were also made. The sum of the squared deviations between these measured values and the values shown in Table I was 0.0067.

To determine the misregistration between any two pairs of channels from Table I, find the fractional pixel value in the table corresponding to the desired pair of channels. The sign of the entry in the table denotes the direction the channel given by the column must be shifted to register it with the row channel. Positive is defined as in the direction of scan and negative as the opposite direction. For example, channel 1 lags channel 2 and channel 2 also leads channel 3.

TABLE I. SKYLAB S-192 SENSOR MISREGISTRATION (PIXELS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	17	19	20
1		-.50	.05	-.45	-.01	-.51	-.14	-.64	-.33	-.83	-.12	-.62	-.07	-.57	.23	.30	-.55
2			+.55	-0.5	.49	-.01	.36	-.14	.17	-.33	.38	-.12	.43	-.07	.73	.80	-.05
3				-.50	-.06	-.56	-.19	-.69	-.38	-.88	-.17	-.67	-.12	-.62	.18	.25	-.60
4					.44	-.06	.31	-.19	.12	- 38	.33	-.17	[illegible]	-.12	.68	.75	-.10
5						-.50	-.13	-.63	-.32	-.82	-.11	-.61	-.06	-.56	.24	.31	-.54
6							.37	-.13	.18	-.32	.39	-.11	.44	-.06	.74	.81	-.04
7								-.50	-.19	-.69	.02	-.48	.07	-.43	.37	.44	-.41
8									.31	-.19	.52	.02	.57	.07	.87	.94	.09
9										-.50	.21	-.29	.26	-.24	.56	.63	-.22
10											.71	.21	.76	.26	1.06	1.13	.28
11												-.50	.05	-.45	.35	.42	-.43
12													.55	.05	.85	.92	.07
13														-.50	.30	.37	-.48
14															.80	.87	.02
17																.07	-.78
19																	-.85
20																	

The test results indicate that the algorithm which has been developed is, in fact, quite accurate. The measurements made on the even and odd numbered high sample rate SDO's yielded the exact results expected. The measurements made on the 17 pairs of channels were consistent among themselves. The standard deviation of each of these estimates over the 10 lines of data which were employed were also quite small (less than .05 pixels). Measurements made on the second set of 10 lines were also consistent with those obtained from the first set of lines. These results indicate that the method is reliable and that considerable confidence can be placed in the results shown in Table I.

An important question to users of Skylab data is: "What effect does misregistration in the original conical data have on the straightened data?" Answers to this question will be pursued during the upcoming month. An even more encompassing problem which will also be considered is the effect of geometric distortion, boundary pixels and field location errors when processing straightened data.

EFFECTS OF CHANNEL-TO-CHANNEL MISREGISTRATION ON CLASSIFICATION ACCURACY AND ON PROPORTION ESTIMATION

The effects of channel-to-channel misregistration of Skylab data on classification accuracy and on proportion estimation were of particular interest in this current phase in the analysis of Skylab data processing techniques. The fact that Skylab data is spatially misregistered has been established. Whether this misregistration is a cause for concern has not been clearly determined. To address this problem a simulation technique was developed to investigate the effects of channel-to-channel misregistration and an experiment designed to implement that technique. What follows is a brief description of the simulation model and an outline of the proposed experiment.

Skylab resolution elements (pixels) were divided into two classes: (1) pure field center pixels and (2) pixels that fall on field borders, i.e., mixture pixels. Figure 1 is a display of two pixels exemplifying each of these two categories. Pixel (2) is a mixture in each channel of 1/2 CROP W and 1/2 CROP O. The variable α_{wi} will be used to designate the mixture proportion of CROP W and α_{oi} the proportion of O in channel i.

Figure 2 illustrates resolution elements affected by a misregistration of 1/2 a pixel in channel 2. Channel-to-channel misregistration affects each pixel category as follows: (a) pure field center pixels can be misregistered but remain field center pixels; (b) field center pixels can be misregistered so those channel(s) out of registration become mixtures of two or more crop types; (c) mixture pixels can be misregistered so channel(s) out of registration represent

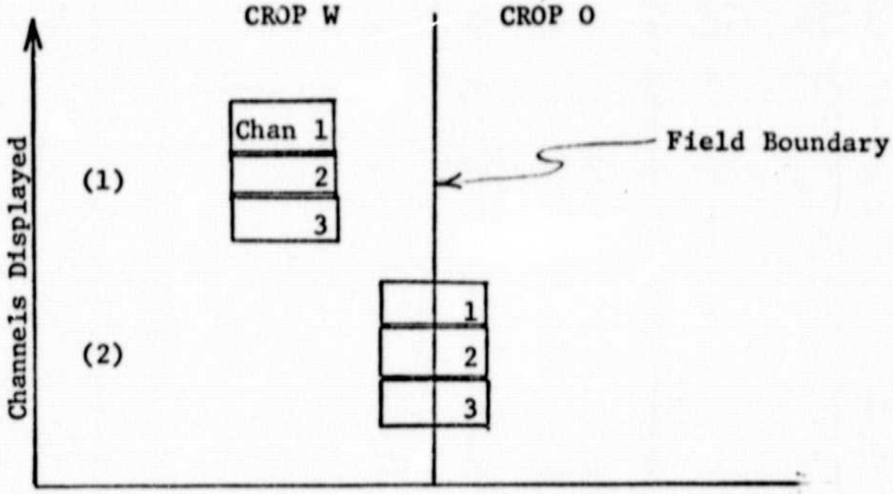

FIGURE 1. ILLUSTRATION OF REGISTERED RESOLUTION ELEMENTS FOR THREE CHANNELS OF DATA

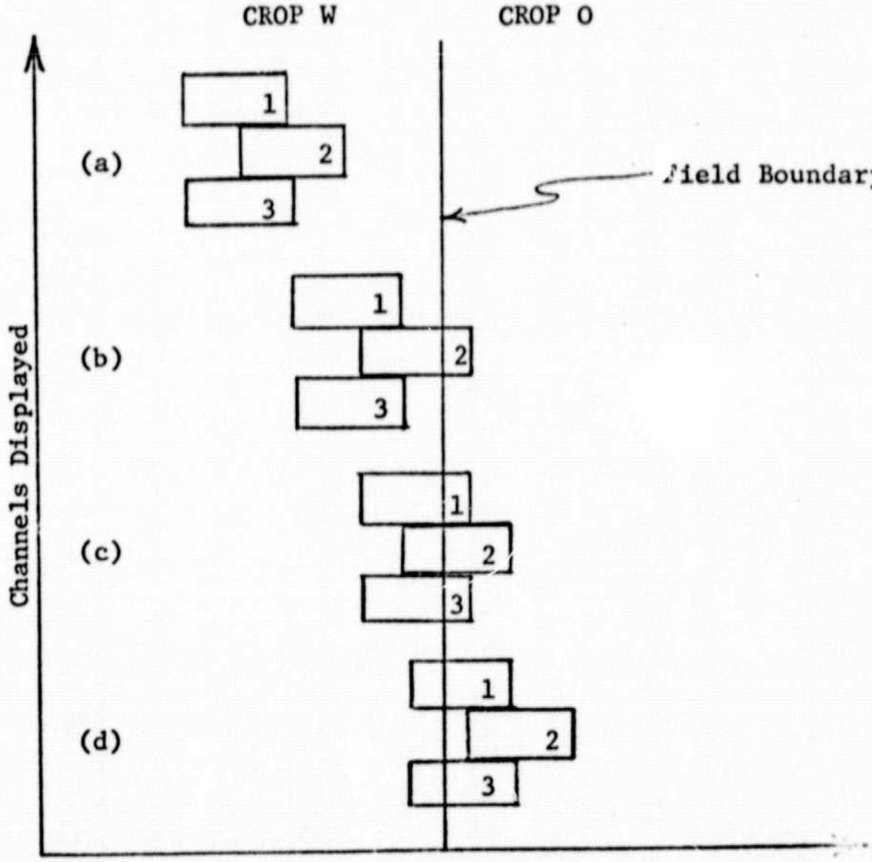

FIGURE 2. ILLUSTRATION OF VARIOUS RESOLUTION ELEMENTS MISREGISTERED ONE-HALF PIXEL IN CHANNEL 2 OF THREE DATA CHANNELS

different mixture proportions; and (d) mixture pixels can be misregistered so those channel(s) out of registration become pure field center values. The variable 'β' will be used to denote the degree of misregistration.

Through statistical analyses of each of the four above categories, two simulation models were developed. The first model simulated the effect of misregistration on pixels that fell into category (a). The second model, a more complex model, summarizes the effects on categories (b), (c) and (d).

Analysis showed that pure field center signatures derived from misregistered data are less correlated in those channels out of registration than field center signatures derived from corresponding registered data. This conclusion was based on an earlier study of correlation.[1] The model chosen to simulate this effect was one that estimated the decorrelation as a linear function of misregistration. That is, given a perfectly registered distribution S_R with means μ_R and covariance C_R, the misregistered representation of this same distribution S_m would have the same means μ_R but different covariance C_m where any term of C_m say c_{mij} is related to a term of C_R in the following manner:

$$c_{mij} = c_{rij} \quad \text{for } i=j$$

$$c_{mij} = c_{rij} \quad \text{for } i \neq j \text{ and } i \text{ and } j \text{ registered with respect to one another, i.e., } \beta=1.$$

$$c_{mij} = \beta c_{rij} \quad \text{for } i \neq j,\ 0 \leq \beta < 1 \text{ and } i,j \text{ misregistered with respect to one another.}$$

where β is dependent linearly on the degree of misregistration.

A model simulating the effect of misregistration on distributions falling in categories (b), (c) and (d) was developed in conjunction with an ongoing SR&T investigation.[2] An added complication was that the correlation term between channels misregistered with respect to one another of pixels representing mixtures had to be carefully determined analytically. Restricting

---

[1] First Quarterly Report, Task IV, "Proportion Estimation", NASA CR-ERIM 109600-3-L, August 30, 1974, H. Horwitz, J. Lewis and A. Pentland.

[2] Prior work is described in: "Studies of Recognition with Multitemporal Remote Sensor Data," NASA CR-ERIM 109600-19-F, Section 3.2 (in publication), W. A. Malila, R. H. Hieber, R. C. Cicone. Spatial misregistration can occur in multitemporal data between sets of channels from the separately acquired data sets.

our effort to misregistration involving mixtures of only two crop types, we found that a component a_m of the mean A_m of a distribution W misregistered into a distribution O is:

$$a_{mi} = \alpha_{wi}\, a_{wi} + (1-\alpha_{wi})a_{oi} \tag{1}$$

where i is the channel and α_{wi} is the proportion of W present in channel i. a_{wi} and a_{oi} are the i^{th} channel means of covers W and O, respectively.

The definition of a term c_{mij} of the variance-covariance matrix of the simulated misregistered distribution is:

$$c_{mij} = \min(\alpha_{wi},\alpha_{wj})*c_{wij} + [1-\max(\alpha_{wi},\alpha_{wj})]*c_{oij} \tag{2}$$

where c_{wij} and c_{oij} are covariance terms for the i^{th} and j^{th} channels of cover W and cover O respectively. If $\alpha_{wi} = \alpha_{wj}$ for all i,j, the model is equivalent to the ERIM Mixtures Simulation Model.

Once simulation models were established, an experiment was designed to aid in the evaluation of the effects of misregistration on field center and mixture pixel classification. A program PEC, developed at ERIM, will be used in the calculation of the expected performance matrix for a given set of signatures input to the program. The program uses a Monte-Carlo type technique to determine the performance matrix which is itself based on a linear boundary classification algorithm. The resultant performance matrix is interpreted under the assumption that the distributions represented by the signatures behave in a Gaussian manner.

The first phase of the experiment involves a study of the effects of misregistration on field center pixels that remain field center in all channels even after misregistration. Five Skylab field center signatures were chosen and the seven best channels are to be used in the analysis. The signature set is assumed to be registered and all simulations established with respect to this initial signature set. Three channels are misregistered in the simulation. This is a parameter that may be varied in future experiments. Simulations of four degrees of misregistration are to be carried out; these are misregistrations of 1/3, 1/2, 2/3 and 1 full pixel. Once the signatures representing each misregistration are calculated, a performance matrix will be produced for each misregistration. Analysis of these matrices should help answer the question: Does channel-to-channel misregistration significantly affect pure field center classification accuracy?

The second phase of the experiment is to study the effects of channel-to-channel misregistration on pixels that are mixture pixels after misregistration in one or more channels. Using the same initial set of signatures previously described, mixtures of all possible pairs of distributions are to be simulated, using the Mixture Simulation Model, for the registered case and for misregistrations of 1/2 and 1 whole pixel. The program PEC will then calculate the expected classification performance of each mixture distribution with respect to the linear decision boundaries between the pure field center signatures. Analysis will consist of the study of the classification curve as a function of the location of the pixel across a field boundary. It will be of particular interest here to examine the false alarm rate of each class of signatures to determine whether misregistration in particular affects this statistic.

EXTRACTION OF SPECTRAL SIGNATURES

A set of spectral signatures were extracted for the major ground covers of the test site. These signatures will be analyzed for discriminability of ground classes, identification of optimum bands for processing this S-192 data set, as inputs to the previously mentioned investigation into misregistration effects, and they will be analyzed to determine the suitability of the signature set for the proportion estimation algorithm. The signatures will be used to classify the test area and also will be used in the signature extension investigation.

The signatures extracted were for 12 bands (SDO 16 was found to be, not only worthless, but a source of confusion in the training procedure and thus was eliminated). The ground covers represented in the signature set were corn, trees, brush, grass, pasture, stubble, water, alfalfa and soybeans. The training procedure used is outlined below.

First, it was necessary to identify field center pixels for each field, that is, those pixels which are sufficiently interior to the field so as to insure that the whole of the area resolved for those pixels lie entirely within the same homogeneous area. Obviously if one wishes to extract a signature for a given class, one must use information from pixels which represent only that class. Identification of field center pixels is accomplished by the inscribing of a smaller similar polygon within the polygon which defines the field being considered. The distance the field center polygon is inset from the original is calculated so that even in the worst case all the pixels in the field center polygon are guaranteed to be resolving only areas within the field.

In general, the inset calculation is a summation of many components, and in fact the inset may be different in the direction of scan than in the along track direction. We can generalize the inset (I: $\{I_x, I_y\}$) as follows:

$$I_\alpha = \left(\frac{D_\alpha}{P_\alpha}\right) B + R_\alpha + L_s + L_c + S$$

where: α indicates x: scan direction or y: line or along track direction

D_α is the size of the resolution cell in the direction of α

P_α is the size of the picture element in the direction of α

B is the inset necessary to insure that the pixel does not include the boundary between fields. Typically B = 0.5 pixel.

R_α is the error due to misregistration effects, e.g., if one channel is misregistered from the others by r pixels, then this channel could still be imaging across the field boundary when the other channels are imaged entirely within the field. For conic data, $R_y = 0$, but for straightened data, in general, $R_x \neq 0$ and $R_y \neq 0$.[1]

L_s and L_c are due to field location errors which may have occurred. L_s is the error in transforming coordinates from the digitized photography to the straightened data. L_c is the error in going from straightened to conic coordinates. In both instances, we used as estimates of L_s and L_c the standard error of Y given by the regression analysis in the calculation of the transformations.

S is the error due to "movement" of individual pixels as a result of the nearest neighbor scan line straightening. For conic data, therefore, S = 0. For straightened data, S = 0.5 pixel.

---

[1]This analysis and subsequent training was carried out before the investigation reported in the first part of this report had been completed. In retrospect, it appears that for conic data, $R_x = .3$, $R_y = 0$; we used $R_x = 0$, $R_y = 0$ assuming that we had correctly and perfectly deskewed the data

The inset we calculated for use in processing the conic data was:

$$I = I_x = I_y = \left(\frac{81}{72}\right) 0.5 + 0 + .52 + .40 + 0 = 1.50 \text{ pixels}$$

by comparison, for processing scan line straightened data we should have used:

$$I = I_x = I_y = \left(\frac{81}{72}\right) 0.5 + 1.0 + .52 + 0 + .5 = 2.6 \text{ pixels}$$

Considering the small size of the fields in the test area, we felt that 1.5 pixels was a very large inset, perhaps leaving an insufficient number of pixels available. Certainly an inset of 2.6 pixels as would be needed for processing straightened data would have excluded our locating field center pixels in this manner. In an effort to see if the calculated inset could be reduced, we thoroughly examined graymaps of the conic data, comparing them to maps of the digitized field locations. It appeared that 0.9 was an excessive number to use as the error in field location, and 0.5 was settled on as being a reasonable upper bound on the location error. Thus we used an inset value of 1.1 in defining field center pixels.

Out of 386 fields originally located in the test area (all fields were bigger than 17 acres), close to 200 had no field center pixels identified and a further 60 had only one field center pixel identified. We were able to use approximately 120 fields with a total number of 1063 field center pixels identified. The total number of pixels in the part of the test site used in this investigation was over 24,000.

Since we suspected that many of the ground cover classes should be represented by more than one spectral signature, instead of combining individual field signatures we generated spectral signatures using a supervised clustering algorithm. Clustering was done using only field center pixels for each ground cover type, and a total of 24 spectral signatures were generated. Three of the signatures were for the village of Williamston. Since these three consist almost entirely of mixture pixels, they were discarded and were not taken into consideration for the rest of the work completed during this reporting period.

The resulting signatures were further examined to determine if any of the signatures, although differently named, were spectrally similar. It was found that some of the pasture and weed signatures were very similar to some of the grass signatures. Since the categories were somewhat nebulous in the first

place it was decided to combine groups of signatures from these classes on the basis of spectral similarity.

During the coming month we intend to begin detailed analyses of these signatures as previously indicated.

Submitted by: Richard F. Nalepka
Richard F. Nalepka
Principal Investigator

Approved by: Jon D. Erickson
Jon D. Erickson
Head, Information Systems and Analysis Department

Approved by: Paul Ri for
Richard R. Legault
Director, Infrared and Optics Division

dd